AD NUMBER

AD833582

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; APR 1968. Other requests shall be referred to Army Behavioral Research Lab., Washington, DC.

AUTHORITY

USABSRL ltr, 31 Jan 1969

AD833582

AD

PROCEDURES FOR ASSIGNING CRITERION GRADES

TO FAILURES AND TURNBACKS

IN ARMY SCHOOL COURSES

Milton H. Maier

MILITARY SELECTION RESEARCH DIVISION

D D C

JUN 1 3 1968

B

BEFRL

U. S. Army
Behavioral Science Research Laboratory

April 1968

23

AD

PROCEDURES FOR ASSIGNING CRITERION GRADES

TO FAILURES AND TURNBACKS

IN ARMY SCHOOL COURSES

Milton H. Maier

MILITARY SELECTION RESEARCH DIVISION
Edmund F. Fuchs, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

Washington, D. C. 20315

April 1968

Wash., D.C. 20315

FOREWORD

The NEW CLASSIFICATION TECHNIQUES Task applies psychological measurement methods to enable the Army to make the best use of the different skills and aptitudes of its enlisted personnel. In a continuing series of studies, research is conducted to attain increasingly accurate and differentiated measures of individual potential. The aptitude area measures are kept up to date and effective by developing new tests and improving existing tests for incorporation into the Army Classification Battery and the Army's aptitude area system.

Currently, a large-scale validation study of experimental and operational tests is being conducted across the full range of the Army's military occupational specialties. The tests are evaluated for their effectiveness in predicting final grades in Army school training courses to which the men are assigned. For varying percentages of cases in the school samples, however, final course grades were not available (in case of failure) or were of questionable interpretation (in case of turnbacks who repeated all or part of the course). The present Technical Research Note deals with the methodological problem of treatment of these cases in order to retain the representativeness of the original samples.

The entire research Task is responsive to special requirements of the Deputy Chief of Staff for Personnel and the U. S. Continental Army Command as well as to requirements to contribute to achievement of the objectives of Department of the Army RDT&E Project 2J024701A722, "Selection and Behavioral Evaluation: Personnel Measurement," FY 1968 Work Program.

J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

PROCEDURES FOR ASSIGNING CRITERION GRADES TO FAILURES AND TURNBACKS IN ARMY SCHOOL COURSES

BRIEF

Requirement:

To determine the procedure to be used in assigning grades to Army school training course failures and turnbacks in the 24,000-man sample for validation of the Army's experimenta' differential classification battery.

Procedure:

Alternative methods of assigning criterion course grades to the academic failures were analyzed for two Army school (MOS) courses to determine which method provided the most appropriate criterion measure. In the case of academic turnbacks, the issue was whether the grades assigned by the school were inflated by the extra training they received.

Findings:

For the failures, no one method showed marked advantage. A single grade was therefore assigned to all failures. This was set et about one standard deviation below the minimum passing mark.

Turnbacks were found to have lower grades than regular passers at given levels of aptitude. No compensatory lowering of their grades for the validation analysis was therefore necessary, and the grades assigned by the schools were therefore accepted. Those for whom no school grades were available were assigned a grade one standard deviation below the mean of regular passers.

Utilization of Findings:

With the procedures selected by these analyses, samples reflecting the full range of Army school training performance could be used in validating tests of the experimental differential classification battery. The biasing influence of excluding an important segment of the sample or of assigning scores to that segment on a judgmental basis was thus avoided.

PROCEDURES FOR ASSIGNING CRITERION GRADE TO FAILURES AND TURNBACKS IN ARMY SCHOOL COURSES

CONTENTS

PROCEDURES FOR ASSIGNING CRITERION GRADES TO FAILURES AND TURNBACKS IN ARMY SCHOOL COURSES

BACKGROUND AND PURPOSE

A large-scale study to validate the Army Differential MOS Battery (ADMOSB) is nearing completion. The tests in the battery are designed to measure knowledge, skills, and interests that are predictive of success in specific occupational fields. Initial selection of tests for the battery was based on validation data obtained on enlisted men within a limited group of military occupational specialties (MOS). The tests showing satisfactory validity were selected for ADMOSB, and the battery was administered to about 25,000 Army enlisted men in over 100 training courses. The experimental tests are to be compared to the operational Army Classification Battery (ACB), and the new tests showing satisfactory differential validity will be incorporated in the revised ACB as replacements for, or supplements to, the operational tests.

The criterion measure to be used in validating the tests is the final course grade earned in the regular training courses. Validity data were collected in a natural operational context, which precluded exercising experimental control over the treatment of the trainees. With so many individuals and training courses, it was inevitable that some individuals would not progress along the normal schedule. About 5,000 men (about 20 percent of the original sample) did not graduate at the scheduled date. About 8 percent of the total sample were academic failures and 6 percent were academic turnbacks who had to repeat portions of a course because of poor performance. The remaining 6 percent withdrew from the courses for nonacademic reasons. The meaning of any course grades reported for failures and turnbacks was questionable; in many cases, grades were not even reported. The grades for the 80 percent that completed the courses on schedule may be considered reasonable indicators of the knowledge and skills the men had at the end of the course. Procedures needed to be developed to estimate criterion grades to be assigned the academic failures and turnbacks that would place them on the same score scale as those who graduated on schedule.

The academic failures are clearly from the low end of the criterion distribution. If they were to be omitted, the distribution of grades would be seriously curtailed. These cases had to be included if the samples were to be representative of the full range of training performance.

End-of-course grades were available for some of the academic failures, but not for all. In developing procedures for assigning criterion grades to all failures, primary consideration was given to maintaining a linear relationship between predictor and criterion. The relationship between predictor and criterion is generally described in

terms of a linear model; the grades assigned to the failures should therefore conform to the model. The decision was made to assign to the failures criterion grades that would maintain the linear relationship observed for the group that completed the courses in the regular time.

The academic turnbacks are also special cases because their instructional time was longer than normal. The primary consideration was whether the grades of the turnbacks should be adjusted to reflect the extra training. If the grades for the turnbacks were found to be higher than for those who graduated at the regular time, holding aptitude constant, then a downward adjustment appeared warranted. If the grades were not higher, then no downward adjustment would be required. Grades of the turnbacks could of course be lowered to compensate for the extra time and cost involved, but such a penalty would be based on administrative considerations, not on the regression of grades on predictor. To determine whether the grades of the turnbacks were inflated, the regression of the turnback grades on a valid predictor was compared to that for the group who completed the course on schedule.

No statistical analyses were performed for the group that withdrew from the courses for nonacademic reasons. In contrast to the failures and turnbacks, no systematic selection factors were operating for this group. Since exclusion of the group would not bias the sample, no attempt was made to find reasonable estimates of the final course grades for the men in this group.

ACADEMIC FAILURES

Alternative Methods of Assigning Grades

Of the 25,000 men originally tested, course grades were obtained for about 24,000. Of this number, 1,990 (about 8 percent) were academic failures. Some were terminal failures, since they completed the course. Some, designated early failures, were dropped before the end of the course. Final grades were available for some of the terminal failures but not for all; no final grades, of course, were available for the early failures.

Three alternative methods of assigning grades to the failures were analyzed. The three methods resulted in three criterion measures:

1. For the terminal failure, the grade reported by the school was used. For the early failures and the terminal failures for whom no grades were reported, an arbitrary score one standard deviation below the minimum passing score was used. The standard deviation was computed on the men who passed the course in regulation time (regular passers). If a terminal failure had a final grade that fell more than two standard deviations below the minimum passing score, the arbitrary decision was

made to move his grade up to two standard deviations below passing. With this method of assigning grades, the distribution of grades was continuous to the point two standard deviations below the passing point, except that the early failures and terminal failures without grades were all bunched at the point one standard deviation below passing.

2. All failures, both early and terminal, were assigned a single grade one standard deviation below the minimum passing grade. Again, the standard deviation was computed on the regular passers. With this method, the distribution of grades was discontinuous between the minimum passing point and one standard deviation below.

3. A distinction was made between early and terminal failures. On the assumption that the terminal failures showed more promise in the course than the early failures, a grade that fell three-fourths of a standard deviation below passing was assigned to the terminal failures, and a grade one and one-fourth of a standard deviation below passing was assigned to the early failures. The early and terminal failures were thereby separated by one-half a standard deviation.

In this third method, the failures were represented by two points, in contrast to a single point in the second method and a more continuous distribution in the first method. The first method made maximal use of the information reported by the school, the second made minimal use of that information by treating all failures alike, and the third made a distinction between early and terminal failures.

Samples

The effects of the three alternative methods of assigning grades to the failures were studied in two training courses, the Airplane Repairman Course (MOS 3511) and the Power Generation Specialist Course (MOS 6661). The Airplane Repairman Course sample had an average failure rate of 10 percent, with 43 failures out of an enrollment of 425. Of the 43 failures, 18 were terminal failures and 25 were early failures. The Power Generation Specialist Course sample had a high failure rate of 22 percent, with 40 failures out of an enrollment of 180. Of the failures, 28 were terminal failures and 12 were early failures.

Statistical Analysis

As a first step in the analysis, product moment correlation coefficients were computed between the Army Classification Battery tests and criterion measures resulting from each of the three methods of assigning course grades. If one method resulted in a more linear relationship, and therefore higher validity coefficients, then that would be the preferred method. The linearity of the regression of grades on a predictor was studied by means of a scatterplot between the General Technical (GT)

Aptitude Area scores and the course grades. The GT score was used because it had a high general validity and the range of GT scores was likely to be large. Both conditions would tend to make any nonlinearity more apparent.

Results

The validity coefficients for the ACB tests and GT against each of the criterion measures are given in Table 1. In the Airplane Repairman course sample, the correlation coefficients of each predictor with the three criterion measures were virtually identical, and the correlation coefficients between the three sets of criterion scores were almost unity. In the Power Generation Specialist Course sample, there were some small differences in the validity coefficients. In general, the validity coefficients for the method in which all failures were assigned a single grade (Criterion B) were one or two points lower than those for the other two criterion measures. The difference, however, was not dramatic enough to make one method of assigning grades clearly superior to the others.

In addition to the correlational analyses, which impose a linear restraint on the relationship, scatterplots between GT and each of the criteria were prepared. If a curvilinear relationship exists, it can be noted in the scatterplot. The regression line for the Airplane Mechanic sample is shown in Figure 1 and for the Power Generation Specialist sample in Figure 2. In both courses, the passing score is 70. The points plotted in the figures are the mean course grades at each level of predictor score. Only one set of means is shown in each figure because in each sample results for the three criterion measures were indistinguishable. With the high degree of similarity in the means, the three regression lines in each sample would virtually coincide.

Since the three methods of assigning course grades to academic failures had similar regression lines, the methods would also have the same degree of linearity. The regression in both samples was essentially linear and the distribution of individual cases around the regression line (not shown in the figures) appeared to be random.

In the absence of any marked differences among the methods, the decision was made to adopt the one computationally most convenient—to assign a single grade to all failures, both early and terminal failures. This grade was set at about one standard deviation below the minimum passing mark. Since the standard deviation of regular passers in most samples was around 5 points, a grade 5 points below passing was assigned to the failures. In a few samples, grades had been assigned on a different scale, and for these samples 12 points (one standard deviation) were subtracted from the passing mark.

Table 1

CORRELATION OF ARMY CLASSIFICATION BATTERY TESTS WITH FINAL COURSE GRADES

Test	Sample									
	Airplane Repairman (MOS 3511) Criterion[a]					Power Generation Specialist (MOS 6861) Criterion[a]				
	A	B	C	Mean	SD	A'	B	C	Mean	SD
Verbal	54	54	54			51	51	52		
Arithmetic Reasoning	56	56	57			60	60	60		
Pattern Analysis	39	38	39			57	54	55		
Mechanical Aptitude	46	46	46			54	54	53		
Clerical Speed	31	31	31			42	41	42		
Radio Code Aptitude	30	30	30			41	39	40		
Shop Mechanics	41	42	41			51	49	51		
Automotive Information	46	45	45			40	39	40		
Electronics Information	40	40	40			55	53	54		
Classification Inventory	32	32	32			27	27	27		
General Information Test	55	55	55			42	41	42		
General Technical Aptitude Area	62	62	62			62	61	62		
Criterion A		99	99 +	77.8	6.9		98	98	74.4	6.4
Criterion B			99 +	77.7	7.1			99	74.3	6.4
Criterion C				77.7	7.0				74.5	6.1

[a]Criterion A - Continuous course grades used for terminal failures
Criterion B - Single grade assigned to all failures
Criterion C - Grade 3/4 S. D. below passing assigned to terminal failures and 5/4 S. D. below passing assigned to early failures

COURSE
GRADES

GT APTITUDE AREA SCORES
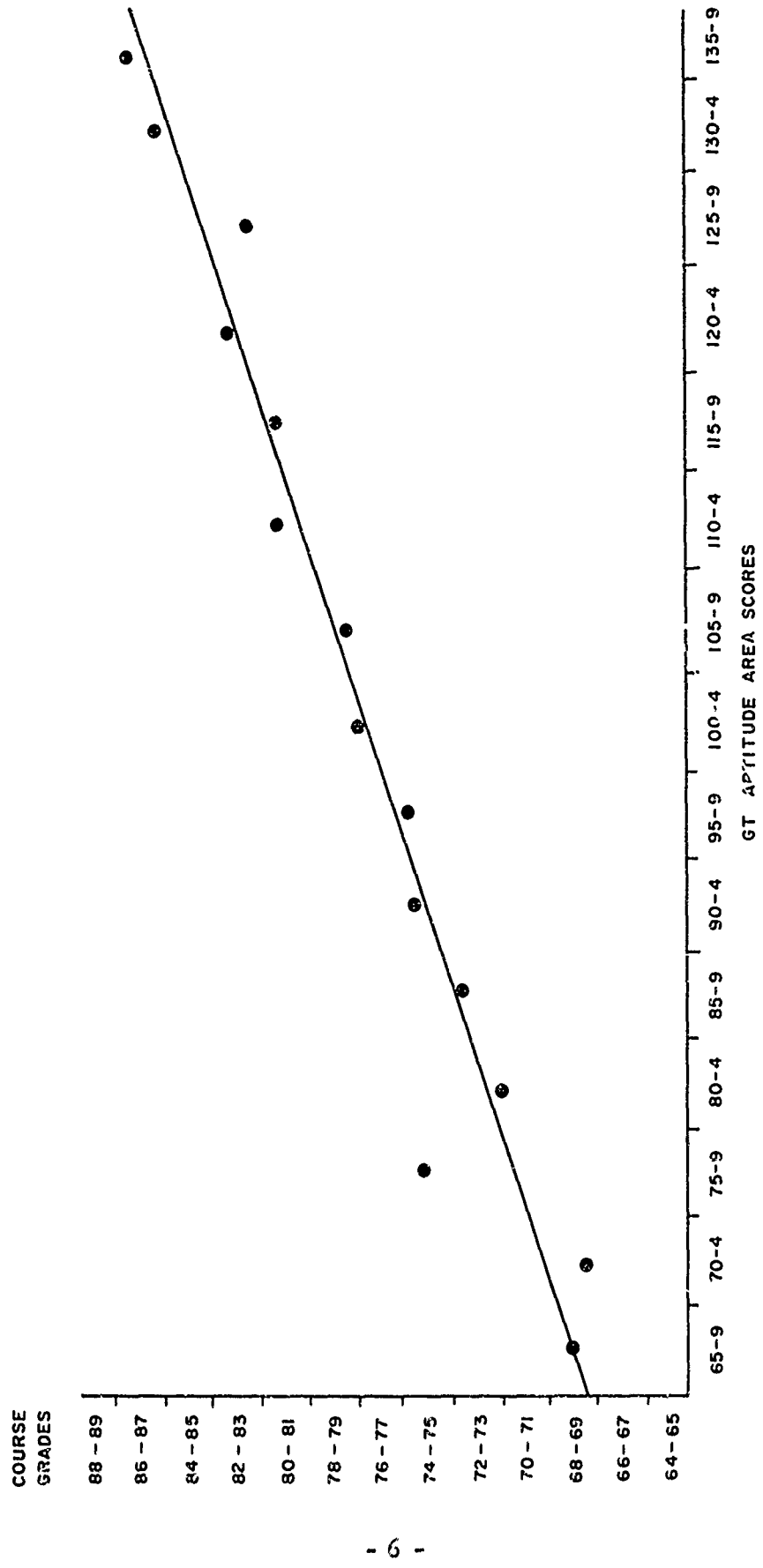
Figure 1. Regression of course grades on General Technical Aptitude Area scores in the Airplane·Repairman (MOS 3511) sample

COURSE
GRADES

82-83
80-81
78-79
76-77
74-75
72-73
70-71
68-69
66-67
64-65
62-63

70-4 75-9 80-4 85-9 90-4 95-9 100-4 105-9 110-4 115-9 120-4 125-9 130-4 135-9 140-4
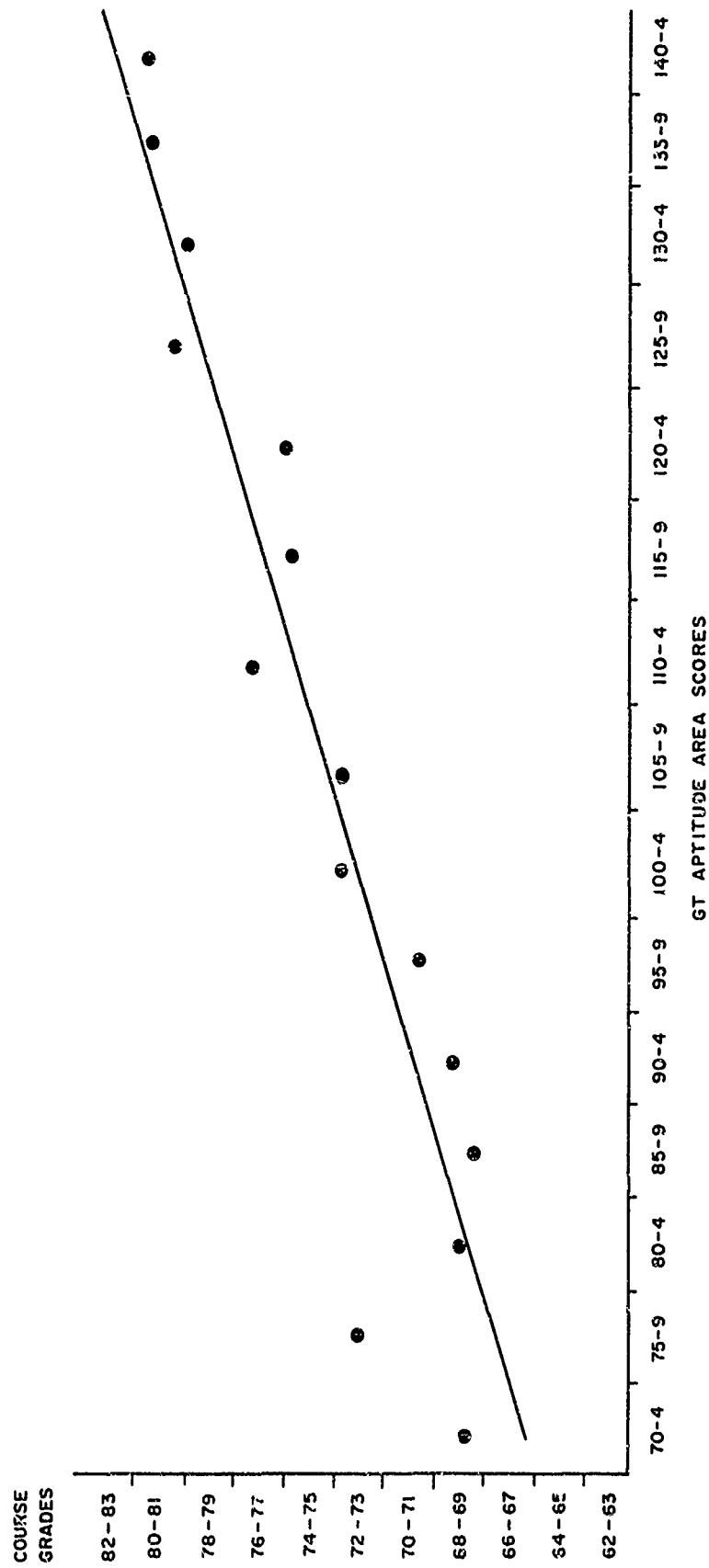
GT APTITUDE AREA SCORES

Figure 2. Regression of course grades on General Technical Aptitude Area scores in the Power Generation Specialist (MOS 6861) sample

- 7 -

ACADEMIC TURNBACKS

Another large group of exceptional cases is composed of the academic turnbacks or recycles; 1395 cases, or about 6 percent, were turnbacks. Men in this group repeated portions of a course because they had done poorly in one or more units of instruction. As mentioned earlier, the issue for the turnbacks was whether to lower their course grades to compensate for the extra instruction they received. The method used to resolve the issue was to compute separate regression equations for turnbacks and regular passers. If the regression line for the turnbacks was significantly higher, then their grades would be adjusted downward to make them comparable to the grades for the regular passers.

Samples

The Arithmetic Reasoning Test (AR) from the ACB was chosen as the predictor because in previous studies it had high validity (r over .60) for similar courses. Two samples were selected for analysis, the Fixed Station Receiver Repairman Course (MOS 2711) and the Radio Relay and Carrier Attendant Course (MOS 2931). The Fixed Station Receiver Repairman sample had a turnback rate of 24 percent, 33 turnbacks out of 141 cases. The Radio Relay and Carrier Attendant sample had a turnback rate of 16 percent, 68 turnbacks out of 420 cases. The regression of course grade on AR was computed separately for passers and academic turnbacks in each sample.

Results

The regression equations for predicting course grades from the Arithmetic Reasoning Test are shown in Table 2. The regression lines for the Fixed Station Receiver Repairman sample are depicted in Figure 3 and for the Radio Relay and Carrier Attendant sample in Figure 4. The regression lines for the academic turnbacks were substantially lower than for the regular passers. In both samples, the predicted course grades for the turnbacks were about one standard deviation below those for the regular passers. For example, in the Fixed Station Receiver Repairman sample, where the standard division of the course grades was 12, an AR score of 100 resulted in a predicted course grade of 97 for the regular passers, but only 84 for the turnbacks. Thus, even though the academic turnbacks received extra instruction during the course, their expected performance was still below that of the regular passers.

The results indicated that no downward adjustment of the course grades was required for turnbacks. The extra instruction did not inflate their grades beyond what could be expected for their level of aptitude; in fact, their expected performance was poorer than for the regular passers. An upward adjustment to make the two regression lines coincide would do violence to common sense. Decision was therefore made to accept the grades for the turnbacks as assigned by the schools.

- 8 -

Table 2

REGRESSION EQUATIONS FOR REGULAR PASSERS, ACADEMIC TURNBACKS, AND TOTAL SAMPLES IN TWO MOS COURSES

Group	N	Course Grades		Arithmetic Reasoning		Regression Equation[a]	Std. Error of Estimate
		Mean	SD	Mean	SD		
Fixed Station Receiver Repairman (MOS 2711)							
Regular Passers	118	101.2	11.8	117.6	11.5	Y = 73.0 + .24X	11.0
Academic Turnbacks	33	87.9	8.6	104.2	13.2	Y = 71.2 + .16X	8.3
Total Sample[b]	194	94.3	14.9	111.9	18.6	Y = 55.0 + .35X	13.3
Radio Relay and Carrier Attendant (MOS 2931)							
Regular Passers	352	85.2	6.0	106.9	17.6	Y = 64.9 + .19X	5.1
Academic Turnbacks	68	77.4	7.3	96.3	16.9	Y = 54.2 + .24X	6.1
Total Sample[b]	482	81.6	8.9	103.2	18.3	Y = 54.0 + .27X	7.1

[a]Y = predicted course grade; X = Arithmetic Reasoning Test score.

[b]Total sample includes academic failures and regular passers and academic turnbacks.
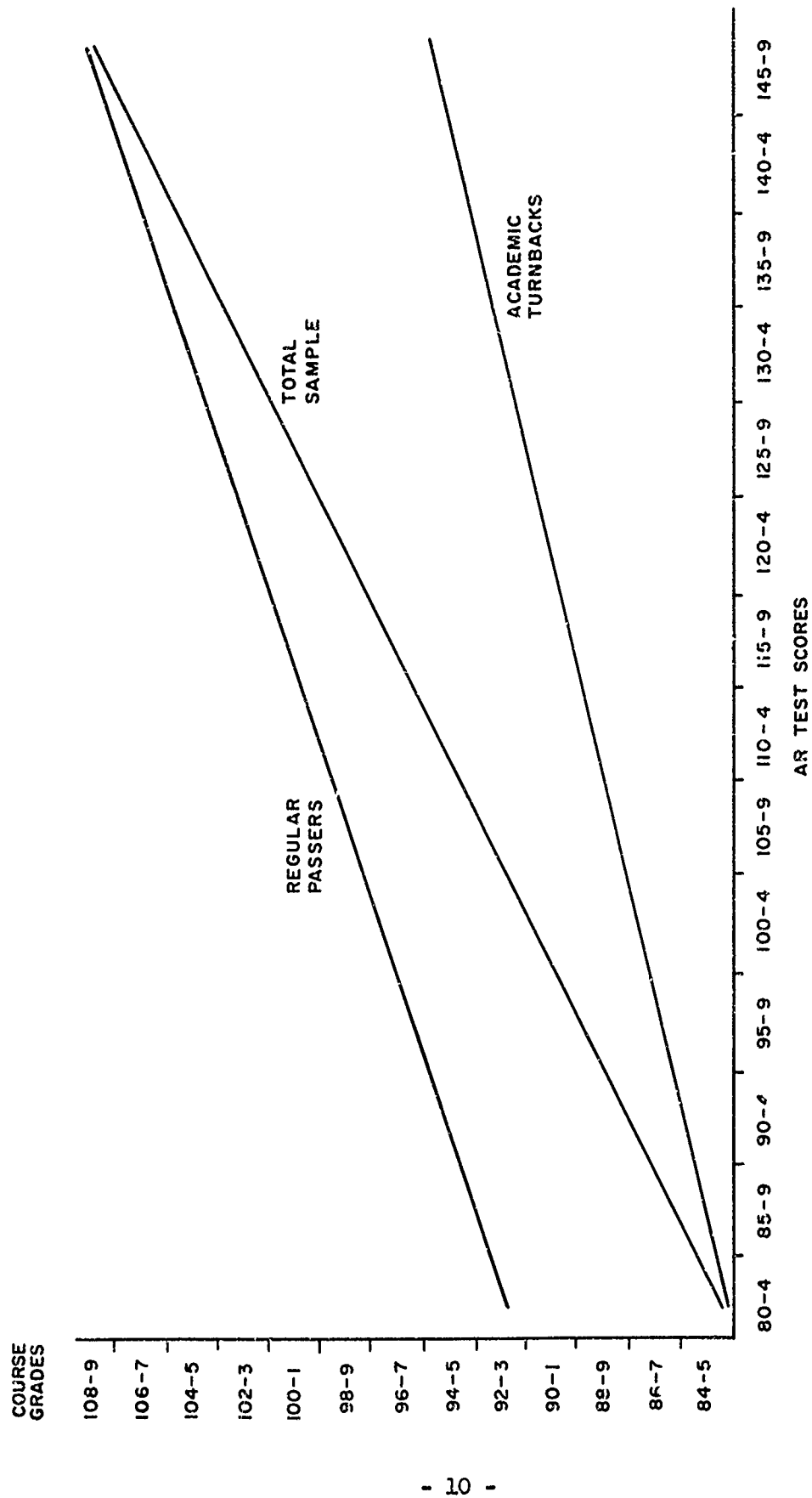
Figure 3. Regression of course grades on Arithmetic Reasoning Test scores in the Fixed Station Receiver Repairman (MOS 2711) sample
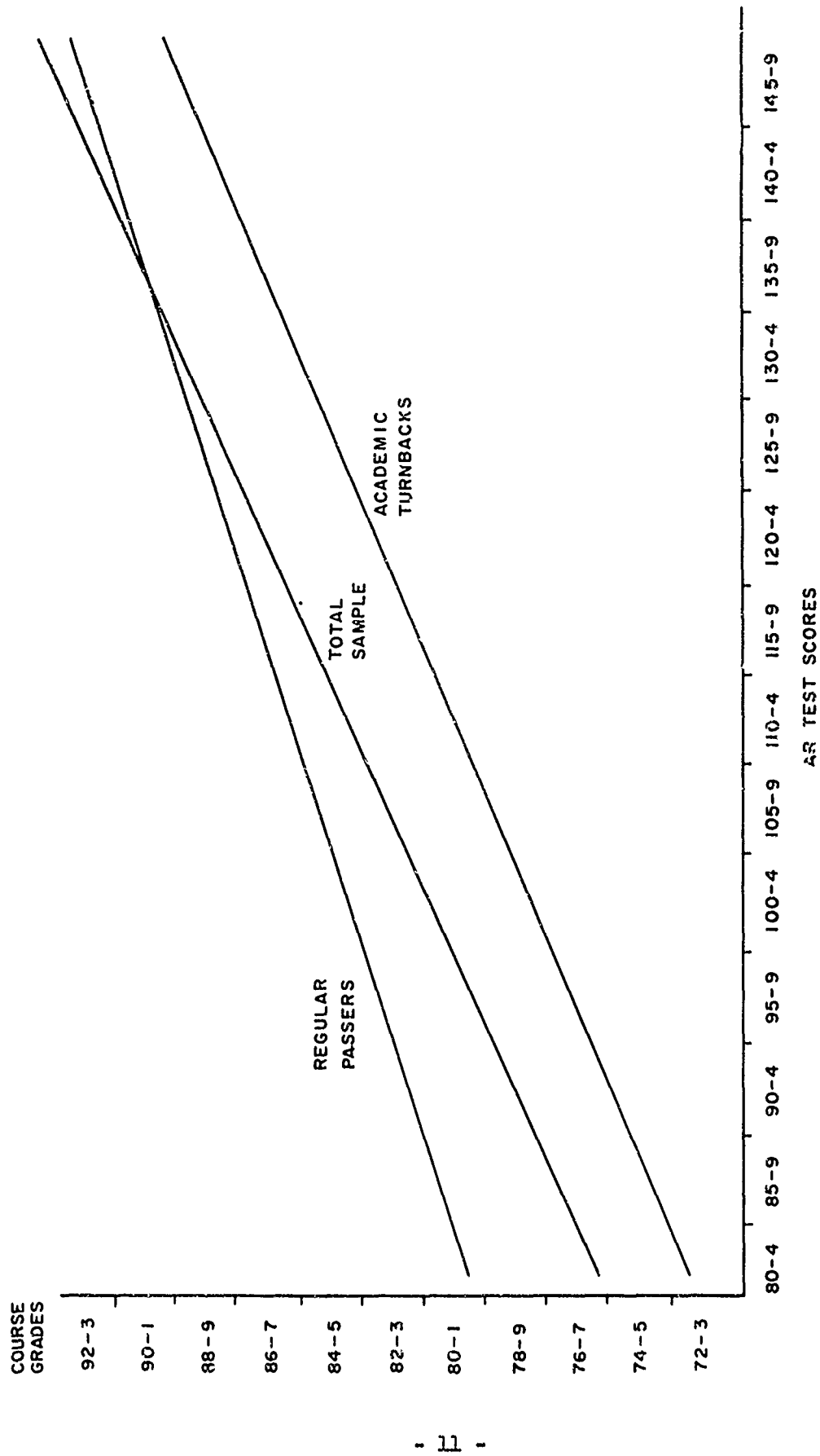
Figure 4. Regression of course grades on the Arithmetic Reasoning Test scores in the Radio Relay and Carrier Attendant (MOS 2931) sample

An effect of pooling the two groups is that the relationship between grades and predictor becomes more curvilinear. The curvilinearity will be such that at the low end of the scale the line will rise more steeply, while at the upper end the line will become flatter. The extent of curvilinearity will depend on the proportion of turnbacks, the difference between predictor means, and the difference between criterion means. In the two samples analyzed, a straight line appeared to fit the predictor-criterion relationship well. Since the percentage of turnbacks in most of the remaining 100-plus courses is smaller than in these two courses, the impact on linearity of regression of pooling the turnbacks and regular passers is expected to be small.

Another implication of pooling the turnbacks and the regular passers is that the standard error of estimate for the pooled group will be increased (see Table 2). This result would mean that the accuracy of prediction for the total sample is not as good as for either subgroup separately. In the present state of knowledge, however, there is no way of predicting before the class starts who will be an academic turnback. Further research is required to find moderator variables that would identify the potential turnback. Until such predictors are found, the reasonable solution is to use a common regression equation for the total sample, even though the errors of prediction are larger than for the separate subgroups.

There remained the question of what grades to assign turnbacks for whom no grades were reported. For each of the 100-plus courses, the mean grades for the regular passers and for the turnbacks with grades were computed. It was found that the means of the turnbacks clustered at a point about one standard deviation below the means of the regular passers. The decision was made, therefore, to assign to the turnbacks without grades a grade one standard deviation below the mean of the regular passers. Separate means and standard deviations for the regular passers were computed for each course. All academic turnbacks could thus be included in the validation studies. Their criterion grades would be those assigned by the school, or, if no grade was reported, the grade that falls one standard deviation below the mean of the regular passers.

SUMMARY

Criterion data were collected for about 24,000 cases in over 100 Army training courses. About 20 percent did not complete the course in the regularly scheduled time. The course grades for academic failures and turnbacks, if available, are of questionable meaning. Because the performance of these groups tends to be below average, they could not be dropped from the analyses without biasing the results.

Regression analyses were conducted to find estimates of scores for academic failures and academic turnbacks. For academic failures, the decision was made to assign scores so as to retain the linearity of regression of grades on a predictor of known validity. It was found that a single score for all failures at the point one standard deviation below the minimum passing score met this requirement. For turnbacks, the issue was whether to lower their grades to compensate for the extra training they received. Results indicated that the grades of the turnbacks were, in fact, lower than for the regular passers. No penalty was imposed on the grades of the turnbacks; they were used as assigned in the courses, or, if no grade was available, they were assigned a score about one standard deviation below the mean of the regular passers.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
U. S. Army Behavioral Science Research Laboratory, OCRD, Washington, D. C.	Unclassified
	2b. GROUP

3. REPORT TITLE

PROCEDURES FOR ASSIGNING CRITERION GRADES TO FAILURES AND TURNBACKS
IN ARMY SCHOOL COURSES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Milton H. Maier

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1968	25	0

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. Army R&D PJ No. 2J024701A.22	Technical Research Note 197
c. New Classification Techniques	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. c-41	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls and each transmittal to foreign government or foreign nationals may be made only with prior approval of the US Army Behavioral Science Research Laboratory, OCRD.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Dep. Chief of Staff, Pers., DA, Wash., D.C. U. S. CONARC, Ft. Monroe, Va.

13. ABSTRACT The NEW CLASSIFICATION TECHNIQUES Task, USA BESRL is conducting a large-scale validation study of experimental and operational tests across the full range of the Army's military occupational specialties (MOS). The experimental tests of the Army Differential MOS Battery (ADMOSB) were administered to about 25,000 enlisted men in over 100 Army school training courses. Final course grades for validation purposes were not available or were questionable for varying percentages of cases in the school samples. The present Technical Research Note deals with the methodological problem of assigning grades to academic failures and turnbacks. Three alternative methods of assigning criterion course grades to academic failures were analyzed for two Army School (MOS) courses to determine which method provided the most appropriate criterion measure. No one method showed a marked advantage for the failures and the decision was made to assign a single score for all failures at a point one standard deviation below the minimum passing score. In the case of academic turnbacks, the primary consideration was whether grades assigned by the schools were inflated by the longer instructional time. Since results indicated that grades of the turnbacks were lower than for the regular passers, no compensatory downward adjustment for the validation analysis was necessary. Grades assigned by the schools for this group were therefore accepted. Those for whom no school grades were available were assigned a score one standard deviation below the mean of the regular passers. By use of the procedures developed on basis of the analyses, samples reflecting the full range of Army School training performance could be used in validating the experimental differential classification battery and biasing effects of excluding segments of the sample or of assigning judgmental scores were avoided.

DD FORM 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Military psychology						
Aptitude areas						
Psychometrics						
Psychological measurement methods						
*Correlational analyses						
*Criterion measures						
*Army Classification Battery tests						
*Regression analyses						

S. GOVERNMENT PRINTING OFFICE 1968 1 ·   )

SUPPLEMENTARY

INFORMATION

AD-833 582 Army Behavioral Science Research Lab., Washington, D. C. Technical research note. Rept. no. PESRL-TRN- 197 Apr 68	No Foreign without approval of Army Behavioral Science Research Lab., Attn: CCRD, Washington, D. C.	No limitation	USABSRL ltr, 31 Jan 69

AD-833 582
Army Behavioral
Science Research
Lab.,
Washington, D. C.

No Foreign without
approval of Army
Behavioral Science
Research Lab.,

USABSRL ltr,
31 Jan 69